



硫酯酶催化机制及结构生物学*

周于聪^{1,2)**} 石 婷²⁾ 芦晨阳³⁾ 刘 昊⁴⁾ 白林泉^{2)**}⁽¹⁾ 浙江万里学院生物与环境学院, 宁波 315100; ⁽²⁾ 上海交通大学生命科学技术学院, 上海 200240;⁽³⁾ 宁波大学海洋学院, 宁波 315832; ⁽⁴⁾ 宁波大学医学部基础医学院, 宁波 315211)

摘要 聚酮类和非核糖体肽类化合物是活性天然产物的重要来源, 更是丰富的临床药物先导资源库。位于聚酮合酶和非核糖体肽合成酶末端的硫酯酶结构域在这些活性天然产物的链释放过程中发挥了重要的底物选择功能, 是其合成后期关键催化酶。本文综述了各类硫酯酶的序列和结构特点, 重点分析了硫酯酶结构和催化功能的对应性。此外, 本文还围绕I型硫酯酶的催化机制, 详细阐述了硫酯酶催化的链释放反应的化学本质, 并总结了硫酯酶催化机制的计算模拟方面的研究进展和局限性, 为硫酯酶的结构和机制解析以及理性设计提供参考。

关键词 聚酮合酶, 非核糖体肽合成酶, 硫酯酶, 蛋白质结构, 催化机制

中图分类号 Q518, Q936

DOI: 10.16476/j.pibb.2024.0288

CSTR: 32369.14.pibb.20240288

聚酮化合物 (polyketides, PKs) 和非核糖体肽类化合物 (non-ribosomal peptides, NRPs) 广泛分布于植物、真菌、细菌中, 常作为小分子次级代谢产物参与营养竞争 (抵抗或杀死生境内的竞争者)、细胞生理调控和信号转导、群体感应等过程^[1-3]。模块化的聚酮合酶 (polyketide synthase, PKS) 和非核糖体肽合成酶 (non-ribosomal peptide synthase, NRPS) 及其流水线式的装配形式, 赋予了这些化合物极其多样的结构和活性, 也为利用组合生物合成手段进行模块或结构域的增减、失活、替换, 来编辑和修饰化合物骨架结构带来了无限可能^[4-7]。硫酯酶 (thioesterase, TE) 催化的链释放是最常见的, 也是研究最为透彻的 PKs 和 NRPs 的链释放模式之一^[8-10]。TE 主要通过外源亲核试剂 (如 H₂O) 亲核进攻完成链的水解或酯基转移反应, 亦或是以分子内氧原子、氮原子或碳原子等亲核试剂发起亲核进攻分别实现大环内酯化 (macrolactonization)、大环内酰胺化 (macrolactamization) 或类克莱森缩合 (Claisen condensation) 等完成链释放反应。参与活性天然产物生物合成的 TE 主要有两大类: a. 融合的 I 型硫酯酶 (type I thioesterases, TEIs), 通常分布在 I 型 cis-AT PKS、trans-AT PKS、NRPS 和真菌

PKS/NRPS 等的末端, 以复合物的形式发挥活性^[11]; b. 游离的 II 型硫酯酶 (TEIIs), 在生物合成过程中催化不完整或未正确延伸的短链错误聚酮中间体的释放, 发挥校正或清理功能^[12-15]。此外, 在聚醚类化合物莫能霉素和南昌霉素等的生物合成过程中, 还存在一类独特的游离 TE, 负责其聚酮链的释放^[16-17]。TE 结构的解析, 揭开了其结构和功能对应性的奥秘, 奠定了 TE 分类和序列特点、底物选择性和催化效率的结构生物学基础, 促进了对天然产物链释放机制的理解, 更好地服务于 TE 的定向改造和理性设计。

TE 存在于所有活细胞中, 通过催化存在于多种细胞底物中的硫酯键的裂解而发挥广泛的重要生物学功能。已有文章系统综述了不同 TE 家族的结构、功能和调控机制^[18]。而针对负责天然产物链释放的 TE 家族, Horsman 等^[19] 也回顾了 PKS 和 NRPS TEIs 的多样性、结构和机制, 提出了 TE 充当了天然产物生物合成加载和释放逻辑门的独特观

* 中国博士后科学基金 (2023M743263) 和浙江省博士后择优资助项目 (ZJ2023065) 资助。

** 通讯联系人。

周于聪 Tel: 18817555727, E-mail: zhouyucong@alumni.sjtu.edu.cn

白林泉 Tel: 021-62932418, E-mail: bailq@sjtu.edu.cn

收稿日期: 2024-07-01, 接受日期: 2024-10-14

点，为研究TE演变和功能预测提供了一个重要的视角。本文重点分析了TE蛋白质结构和催化功能的对应性，阐明了TE分类、底物选择性和催化效率的结构生物学基础，促进对天然产物链释放机制的理解。此外，随着计算理论和方法的进步，酶催化反应的理论计算研究大力支持了酶的机制解析及其理性设计。本文也详细综述了TEIs催化的链释放反应进程，及其计算模拟方面的研究进展和局限性。本文将为TE结构和机制解析，以及TE的理性设计提供参考。

1 TE序列特点和蛋白质结构

1.1 TE的进化和序列特点

TE序列进化聚类受其分类（TEI或TEII），或所催化的化合物结构特点，亦或是生物合成酶所属类别（PKS、NRPS、真菌NRPS/PKS等）的影响，

与其催化的链释放化学反应特性无关。TEI的底物选择性（上载选择性和释放选择性，如环化、水解等）无法用典型的隐马尔科夫模型（Hidden Markov models, HMM）或保守基序（motif）高效预测。TE蛋白序列进化聚类（图1）显示，TEI和TEII隶属于独立的进化起源，来自同一种活性天然产物生物合成基因簇的TEI和TEII分属于不同的进化分支，如红霉素生物合成相关的EryTII（TEII）和聚酮类化合物红霉素TE（DEBS TE）（TEI）、PimI（匹马霉素生物合成相关TEII）和PimTE（TEI）等。Wang等^[12]相对系统地比对和总结了TEII的进化分类，主要分为类氨酰转移酶类TEII（aminoacyltransferase-like TEII，如WS20^[20]等）、中间体释放类TEII（intermediate releasing TEII，如RedJ等）和编辑类TEII（editing TEII，如RifR等）。

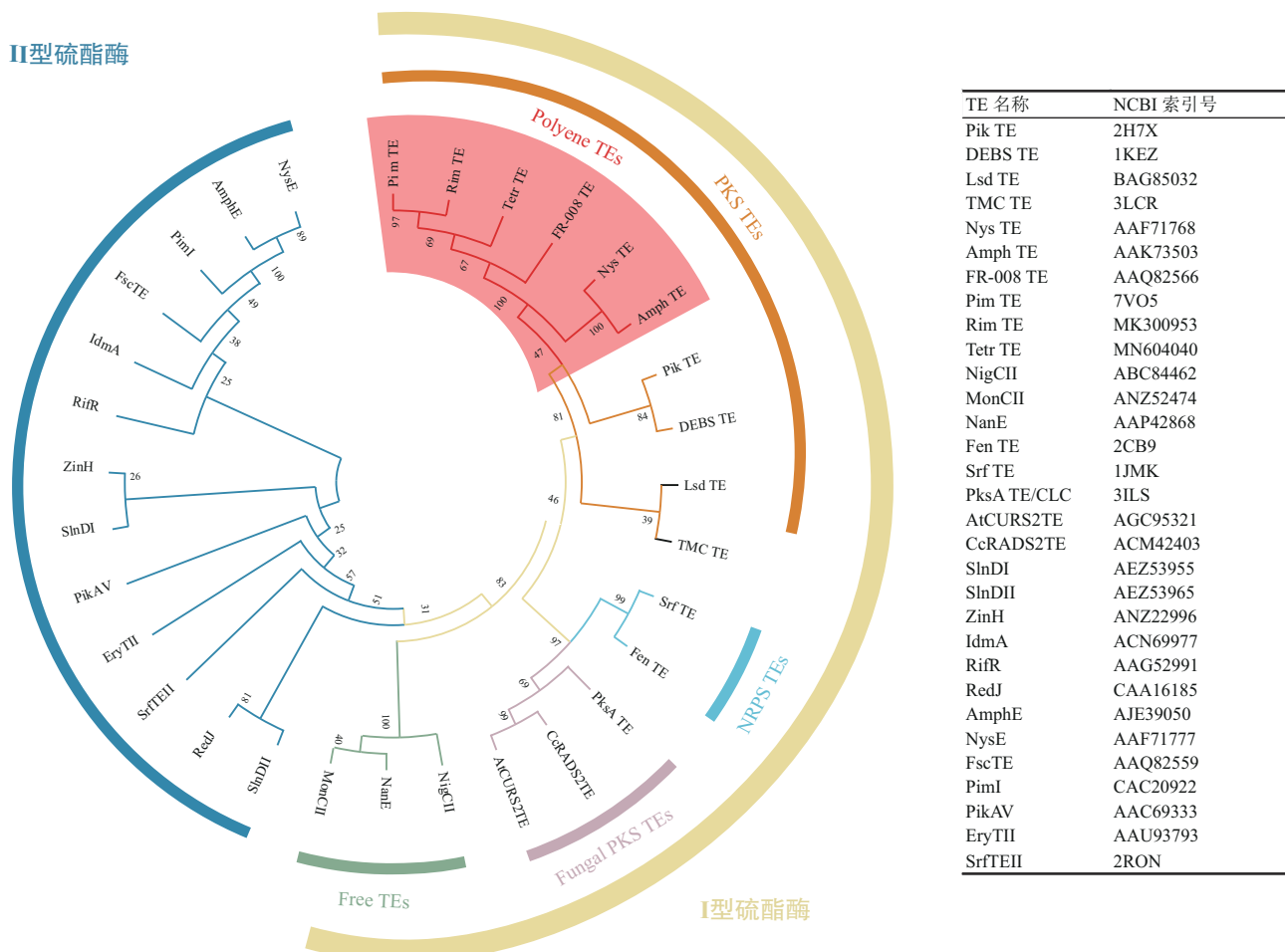


Fig. 1 Phylogenetic tree analysis on selected thioesterases

图1 典型硫酯酶系统进化树分析

硫酯酶的系统进化分析采用MEGA-X软件最大似然法原理（maximal likelihood），bootstrap=500评估校验。Polyene TEs：多烯类抗生素硫酯酶结构域；PKS TEs：聚酮合成酶硫酯酶结构域；NRPS TEs：非核糖体肽合成酶硫酯酶结构域；Fungal PKS TEs：真菌聚酮合成酶硫酯酶结构域；Free TEs：游离硫酯酶结构域。

虽然 TEI 序列非常多样化, 但 TEI 的蛋白质序列聚类分支与其催化的活性天然产物化学结构及其隶属的生物合成酶类型显著相关。Horsman 等^[19]利用 ClustalW 分析了近 300 个 TE 氨基酸序列发现, 其序列中很难具有连续 5 个氨基酸残基同源性高于 50% 的情况。从蛋白质序列聚类结果看, TEI 主要被聚类为 PKS TEI、NRPS TEI、真菌 PKS TEI 和游离 TEI 等分支。真菌来源的 TEIs 在进化起源上与细菌 NRPS TEI 更相近。特别地, NanE、MonCII、NigCII 等来自聚醚类小分子生物合成途径的硫酯酶, 明显聚类在 TEI 大分支, 但又独立于与其他作为结构域嵌合在生物合成酶合酶内部的 PKS TEI 和 NRPS TEI。因 NanE 等游离于 PKS 合酶外, 目前很多文献和综述中都将 NanE 等这种负责聚醚链释放

的酶归为 TEII。但从蛋白质序列聚类结果、蛋白质拓扑结构、催化三联体 Asp 残基的位置以及负责催化成熟的长链聚醚骨架链释放等多方面分析, NanE 等“环氧水解酶”应该被归为 TEI。

1.2 TE蛋白结构

DEBS TE^[21]、苦霉素 Pik TE^[22] 和非核糖体肽类化合物脂肽 TE Srf TE^[23] 的蛋白质结构是最先被研究报道的。截至 2024 年, 已有近 20 种参与活性天然产物合成加工的 TEI 和 TEII 结构被解析 (表 1)。其中, DEBS TE、Pik TE 和 Pim TE^[24] 负责 12 元、14 元或 26 元大环内酯的环化释放。而来自变构菌素 (Tautomycin) 生物合成途径的 TMC TE 是第一个结构被解析的催化聚酮链水解释放的 TE^[25]。

Table 1 High-resolution structures of TEs
表1 硫酯酶晶体结构信息汇总

类型	菌株	化合物	TE名称	PDB索引号	发表年份	参考文献	
I型	聚酮合	<i>Saccharopolyspora erythraea</i>	Erythromycin	DEBS TE	1KEZ, 1MO2	2001	[21]
硫酯酶	酶I型硫酯酶	<i>Streptomyces venezuelae</i>	Pikromycin	Pik TE	1MNQ, 1MN6, 1MNA, 2H7X, 2H7Y, 2HFK, 2HFJ	2003	[22, 26]
		<i>Streptomyces</i> sp. CK4412	Tautomycin	TMC TE	3LCR	2010	[25]
		<i>Streptomyces chattanoogensis</i>	Pimaricin	Pim TE	7VO4, 7VO5	2021	[24]
	非核糖	<i>Bacillus subtilis</i>	Surfactin	Srf TE	1JMK	2001	[23]
	体肽合	<i>Bacillus subtilis</i>	Fengycin	Fen TE	2CB9, 2CBG	2006	[27]
	成酶I型	<i>Nocardia uniformis</i> subsp.	Nocardicins	NocTE	6OJC, 6OJD	2019	[28]
	硫酯酶	<i>tsuyamanensis</i>					
		<i>Streptomyces tsusimaensis</i>	Valinomycin	Vlm TE	6ECE, 6ECD, 6ECF, 6ECC, 6ECB	2018	[29]
		<i>Escherichia coli</i> K-12	Enterobactin	EntF	2ROQ, 3TEJ	2008	[30-31]
	真菌I型	<i>Aspergillus parasiticus</i>	Aflatoxin	PksA TE	3ILS	2009	[32]
硫酯酶		<i>Moorea producens</i> 19L	Curacin	CurM	3QIT	2011	[8]
		<i>Escherichia coli</i>	Colibactin	ClbQ	5UGZ	2017	[33]
	II型硫酯酶	<i>Bacillus subtilis</i>	Surfactin	SrfTEII	2K2Q, 2RON	2008	[34]
		<i>Streptomyces coelicolor</i>	Prodiginine	RedJ	3QMW, 3QMV	2011	[35]
		<i>Streptomyces</i> sp. WAC02707	Borrelidin	BorB	6VAP	2020	[36]
		<i>Streptomyces</i> sp. SNM55	Cyclodepsipeptide	WS5	7E3Z	2021	[20]
			WS9326A				
		<i>Streptomyces actuosus</i>	Nosiheptide	NosK	5V7O	2017	[37]
		<i>Amycolatopsis mediterranei</i>	Rifamycin	RifR	3FLA, 3FLB	2009	[38]
		<i>Escherichia coli</i>	Yersiniabactin	YbtT	6BA8, 6BA9	2017	[39]

上述已解析 TE 蛋白质结构显示, 它们共享类似的三维堆叠结构, 是典型的 α/β 水解酶超家族蛋白^[18]。由一连串 β 折叠和 α 螺旋交替折叠形成疏水核心, 保守活性中心催化三联体 Ser-Asp-His 位于 α/β 疏水核心的顶部, 被一个类似“盖子区域”

(lid region) 的结构包裹 (图 2)。TE 的结构解析促进了 TE 二级拓扑结构分析 (图 3), 揭示了催化活性天然产物链释放的 TE 的空间结构共性和差异, 为 TE 分类和功能预测奠定了结构生物学基础。

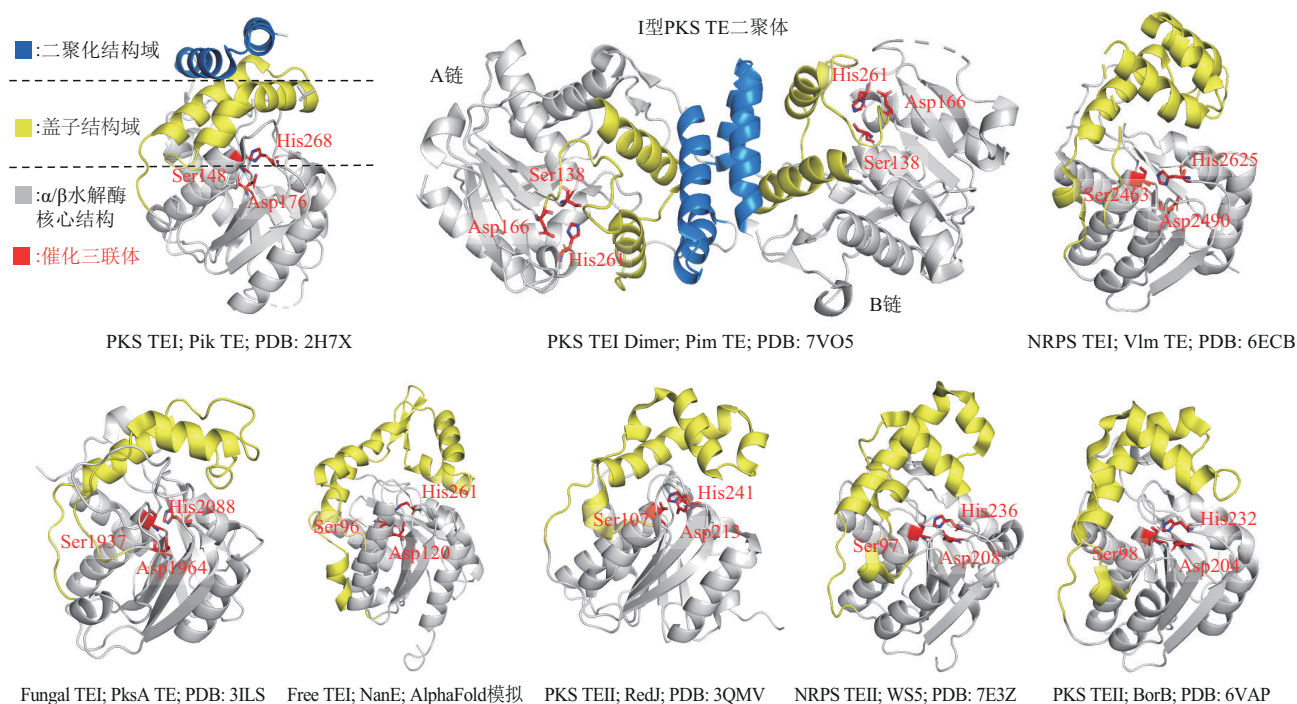


Fig. 2 Structural diagrams of typical thioesterases

图2 典型硫酯酶蛋白三维结构

二聚化结构域以蓝色显示；硫酯酶独特的“盖子结构”（lid region）以黄色显示； α/β 水解酶疏水核心结构以灰色显示；活性中心催化三联体以红色短棍显示；硫酯酶分类、名字及其PDB数据库检索号显示于相应蛋白质结构的下方。PKS TEI：I型聚酮合成酶硫酯酶结构域；PKS TEI Dimer：I型聚酮合成酶硫酯酶结构域二聚体；NRPS TEI：I型非核糖体肽合成酶硫酯酶结构域；Fungal PKS TEI：I型真菌聚酮合成酶硫酯酶结构域；Free TEI：游离I型硫酯酶结构域。PKS TEII：II型聚酮合成酶硫酯酶结构域；NRPS TEII：II型非核糖体肽合成酶硫酯酶结构域。

1.2.1 TE N端结构与二聚化

TEII和NRPS TEIs N端通常缺失 β 1折叠结构，游离的TEI则保留了完整的 β 1折叠和 β 2折叠（图3），它们均以单体的形式发挥催化活性。而PKS TEIs则在 β 1折叠N端还存在两个额外的 α 螺旋（ α I和 α II）结构（图2），以此为基础形成二聚体行使功能，这种结构设计也与I型聚酮合成酶整体二聚化结构契合。

已有的PKS TEI结构，DEBS TE、Pik TE、TMC TE和Pim TE等显示其二聚体结构主要来自 α I和 α II的疏水氨基酸残基通过疏水相互作用维系（图4）。DEBS疏水二聚体面主要由4个亮氨酸（Leu）残基，Leu18、Leu37、Leu38、Leu41，和1个苯丙氨酸（Phe）残基，Phe44，组成^[21]。TMC TE二聚体面疏水氨基酸组成与DEBS TE类似，由Phe6、Leu9、Leu26、Leu32等残基组成。Pik TE则是由4个Phe残基（Phe12、Phe16、Phe28和Phe38）组成^[22]。同样地，Pim TE单晶结构表明，Pim TE二聚体同样也由疏水相互作用维系，关键

疏水氨基酸为Leu12、Met15、Met31和Leu35^[24]。不同PKS TEI的疏水二聚体面的氨基酸组成和残基位置并不保守，但主要残基均为Leu、Phe、缬氨酸（Val）等疏水残基。在PKS TEIs中，N端这两个二聚体 α 螺旋结构不仅部分参与底物通道的组成，还靠近酰基载体蛋白（acyl carrier protein, ACP）结合位点^[21]。这种性质相似但残基不保守的配置可以确保来自不同途径的两个TE单体以恰当的角度形成二聚体和合适的底物通道结构，促进ACP结构域的正确结合和底物的高效释放。

1.2.2 活性中心催化三联体

作为 α/β 水解酶家族成员，TE也具有典型的丝氨酸（Ser）-天冬氨酸（Asp）-组氨酸（His）催化三联体结构（图2）。其中Ser位于 β 5折叠和 α C螺旋连接的拐角处，负责亲核进攻肽基载体蛋白（peptidyl carrier protein, PCP）/ACP-磷酸泛酰巯基乙胺臂（phosphopantetheine, Ppant）与底物之间的硫酯键，随后和聚酮中间体形成酰基化中间体acyl-O-TE。个别TE活性中心的Ser残基被半胱氨

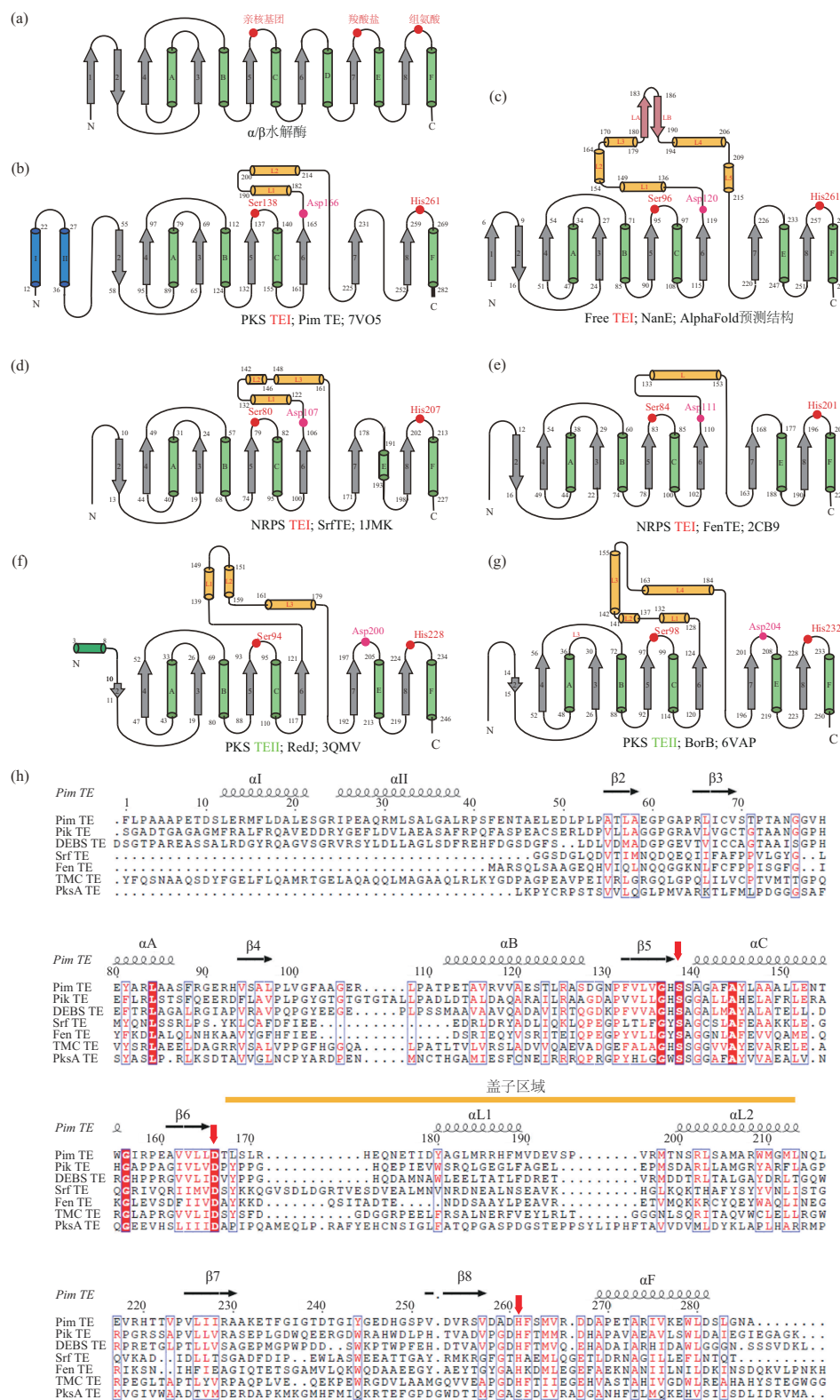
Fig. 3 Schematic diagram of the topological structure of the α/β hydrolyase family of enzymes and thioesterases

图3 典型的硫酯酶二级结构拓扑图

(a) α/β 水解酶超家族蛋白二级结构模式拓扑图; (b) PKS TEI匹马毒素硫酯酶Pim TE拓扑结构; (c) 游离I型硫酯酶NanE蛋白二级结构拓扑图, NanE结构为AlphaFold预测结果; (d, e) 非核糖体肽类化合物Surfactin和Fengycin生物合成相关I型硫酯酶Srf TE和Fen TE蛋白二级结构拓扑图。灰色箭头表示 β 折叠, 绿色圆柱体为 α 螺旋, 黑线为连接各个二级结构的loop结构。黄色圆柱体和粉色箭头特指硫酯酶盖子区域中的重要二级结构, 蓝色圆柱体为PKS硫酯酶特有的形成二聚化的N端 α 螺旋结构。(h) 常见I型硫酯酶的序列比对及其对应的二级结构示意图。 α : α 螺旋 (α -helix), β : β 折叠 (β -sheet)。红色箭头所指残基为活性中心催化三联体残基Ser-His-Asp。

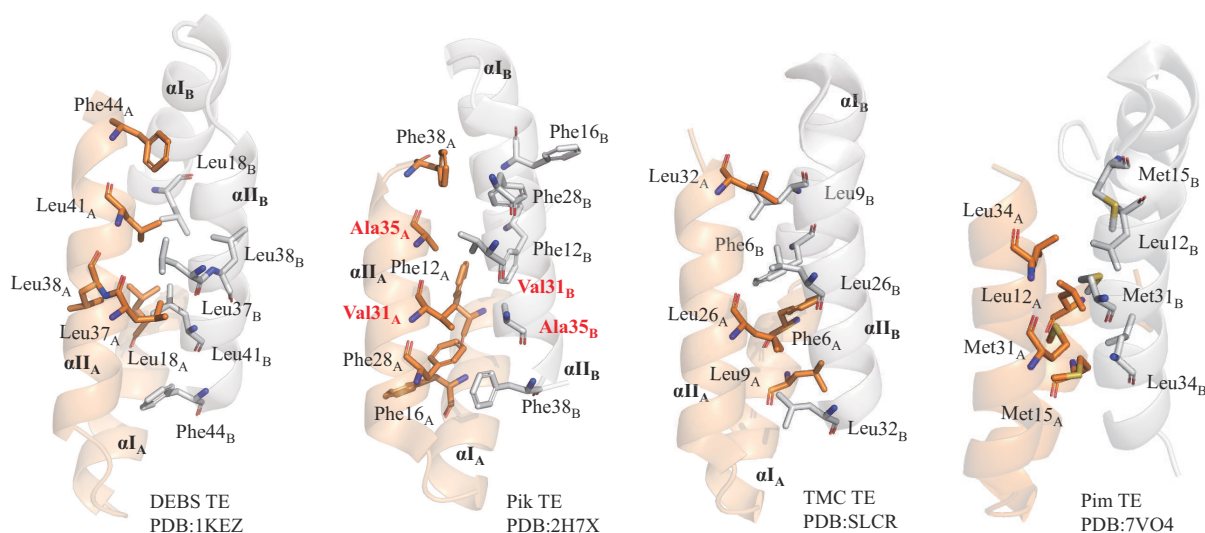


Fig. 4 Hydrophobic dimer interactions in DEBS TE, Pik TE, TMC TE, or Pim TE

图4 DEBS TE、Pik TE、TMC TE和Pim TE疏水二聚体面的氨基酸组成

酸 (Cys) 取代, 如催化多黏菌素 (polymyxin, PMB) TE^[40] 和 Obafluorin (Obi) TE^[41] 环化释放的 NRPS TEIs。组氨酸残基 (His) 位于最后一个 α F 螺旋的顶部, 在 DEBS TE、Pik TE 和 Pim TE 等中 α F 螺旋靠前部分存在二级结构的扭转 (图2), 使 His 靠近活性中心辅助 Ser 或末端羟基的去质子化, 增强 Ser 或末端羟基亲核进攻的活性。而 Asp 在催化三联体中主要发挥增强 His 拔质子活性的作用。Asp 的位置在 TEI 和 TEII 中存在显著区别, 是 TE 分类的重要标准之一。TEI 中 Asp 常位于 β 6 折叠尾部, 其后紧跟着盖子区域; 而在 TEII 中, Asp 则位于盖子区域后, β 6 折叠和 α D 螺旋形成的转角顶部。

1.2.3 盖子区域和底物结合口袋

不论是 TEI 还是 TEII, 其二级结构中, 均在 β 6 和 β 7 之间插入了一个盖子区域 (图3)。这是 TE 底物结合口袋的重要组成部分, 也是 TE 与 α/β 水解酶家族蛋白模式结构相比最显著和最重要的二级结构差异。这个盖子区域由 α/β 水解酶模式结构的 α D 螺旋变化而来, 通常由数个 α 螺旋和 loop 组成。组成盖子区域的氨基酸序列保守性低, 且序列长度也不尽相同, 折叠的结构非常多样化, 是识别和催化 (环化或水解) 不同结构的非核糖体肽中间体或聚酮中间体的重要结构基础。DEBS TE、Pik TE 和 Pim TE 等的盖子区域长约 50 aa, NRPS Srf TE 长约 63 aa, NRPS Fen TE 长约 51 aa。特别地, 根据

AlphaFold 建模结果, 游离 TEI NanE 预测结构中盖子区域序列长度近 80 aa, 二级结构也非常复杂, 由 5 个 α 螺旋和 2 个 β 折叠组成。

盖子区域中 α 螺旋数量和螺旋程度可以有效调控底物结合口袋构象变化和空间大小, 与底物相互作用密切相关。NRPS Srf TE 因不含 N 端二聚化的螺旋结构, 其盖子区域主要堆积形成了一个类似“碗”状的底物结合口袋, 且构象变化空间相对较大, 存在明显的“开”和“闭”构象转变^[23]。而 NRPS Fen TE 结构中盖子区域折叠形成了长约 25 Å、深约 6 Å 的“峡谷”结构, 活性中心催化三联体 Ser-Asp-His 坐落于“峡谷”底部。Fen TE 的盖子区域较 Srf TE 更短, 活性中心 Ser 始终暴露在盖子区域形成的狭缝中, 因此并不限制底物进入底物口袋上载到 Ser 残基上^[27]。但 Fen TE 与苯甲基磺酰氟 (phenylmethanesulfonyl fluoride, PMSF) 复合物结构以及 Fen TE 与底物的复合物模拟结果显示, Fen TE 催化 Fengycin 环化释放时, 底物末端残基与活性中心邻近残基的相互作用以及氧负离子洞的形成十分关键^[27]。Liu 等^[30] 解析了首个 NRPS 合酶中 PCP 结构域和 TE 结构域的复合物结构 (EntF, PDB: 3TEJ), 揭示盖子区域还参与了 PCP 结构域的结合。EntF 的盖子区域由两个长的 α 螺旋 (α L1 和 α L2) 组成, 两个 α 螺旋相互平行, 其右侧是 PCP 结构域的结合位点。PCP 结构域中修饰的 Ppant 分子对接在活性中心右侧, 盖子区域底

部和 α/β 水解酶疏水核心结构形成的裂缝中。因此, Liu等^[30]推测活性中心左侧空腔为底物分子结合位点, 由盖子区域左侧loop区(活性中心Asp和 α L1之间的loop序列)和底部 β 6和 α E或 β 7之间的loop组成。不同NRPS TEI结构中, 这两段组成底物结合口袋的序列二级结构有所差异, 部分序列会形成 α 螺旋以改变底物口袋的空间构象。

PKS TE中, 盖子区域和顶部二聚化的 α 螺旋组成了一个两端开口的底物通道结构。通道入口一侧为盖子区域中两个 α 螺旋的转角loop区, 活性中心位于底物通道中心位置。DEBS TE和Pik TE的底物通道相对平直, 通道直径较大, 通道的尺寸会随着溶液环境pH的变化而变化^[22]。匹马霉素硫酯酶Pim TE共晶结构揭示了负责多烯两性分子释放的TE中独特的“双漏斗”底物通道结构^[24]。其活性中心催化三联体所在区域十分狭窄, 保证了底物在预环化构象组织时, 末端羟基被限制在活性中心出口一侧, 促进底物的高效环化, 抑制其他水解产物的积累。多烯硫酯酶盖子区域氨基酸亲疏水性分布也十分契合多烯分子共轭双键疏水区和多元醇亲水区的两性化学结构特点^[24]。此外, 同源建模结果显示, 盖子区域的亲、疏水氨基酸分布还与其释放的产物内酯环大小可能相关。多烯硫酯酶可能通过延长盖子区域中loop长度和降低 α 螺旋化程度, 来构建更大空间的底物通道以释放更大环的聚酮内酯产物^[24]。

针对游离TEI, 目前尚未有相关蛋白质晶体结构解析和报道。而游离TEI大多在聚醚类聚酮活性天然产物的生物合成途径中被发现, 如南昌霉素NanE、莫能霉素MonCII、尼日利亚菌素NigCII、聚醚离子载体J1-001-2 JenE等蛋白质。AlphaFold计算建模和氨基酸序列比对结果显示, 这类游离TEI的盖子区域不仅序列长, 而且二级结构复杂, 可能还罕见地存在 β 折叠结构(图3), 这可能与识别和催化长链聚醚聚酮链的释放密切相关。

TEII作为另一类游离于PKS合酶或NPRS合酶外的独立的TE, 通常在聚酮类和非核糖体肽类天然产物生物合成过程中发挥纠错功能。失活TEII会引起聚酮类和非核糖体肽类天然产物的产量显著下降^[12-13]。最新TEII RifR、RedJ、BorB、NosK、YbtT和WS5等蛋白质晶体结构研究显示, 其盖子区域主要由3~4个 α 螺旋组成, 形成的底物结合口袋相对较浅, 且活性中心充分暴露, 易被结合。

RifR的盖子区域中 α L1螺旋可能控制了底物能否接近活性中心^[38]。这种浅且开放的底物口袋构象使得TEII的底物选择性较宽泛, 可以非特异地释放错误延伸单元, 以发挥纠错功能。研究表明, TEII可能存在多种构象, 但只有一种构象能与携带天然延伸中间体的PCP结构域相互作用^[34]。

1.2.4 “ α E”螺旋

不同TE结构中还存在显著差异的“ α E”螺旋。NRPS TE结构中, “ α E”螺旋长短不一, 且 β 6与 α E之间的loop序列可能参与了底物结合口袋的组成^[23, 27]。PKS TE中, DEBS TE缺乏这段“ α E”螺旋, β 6和 β 7之间完全由一段长的柔性loop组成, Pik TE中含一段极短的4个氨基酸残基组成的“ α E”螺旋。Pik TE的底物口袋以两端贯通的通道结构, 这段 β 6和 β 7之间loop或loop-“ α E”-loop结构不直接参与底物通道的构成, 但可能起到支撑和稳定盖子区域的作用^[22]。

2 TEI催化机制

2.1 TE催化的化学反应本质

TE催化的链释放反应, 其化学本质为两步亲核进攻反应, 分别对应底物上载和产物释放(offloading/release)两个过程(图5)^[18, 42]。底物上载指聚酮或非核糖体肽等中间体在完成骨架链延伸后, 从最后一个模块的ACP/PCP上转移到TE活性中心Ser, 形成乙酰硫酯(肽酰硫酯)的过程。该过程由催化三联体中的His发起, 其在Asp的活化下, 夺取Ser侧链羟基上的质子。去质子化的Ser侧链羟基氧原子亲核进攻ACP/PCP-Ppant, 与底物之间的乙酰硫酯键或肽酰硫酯键的羰基碳原子(C1)形成(peptidyl)acyl-O-Ser复合物。在亲核进攻过程中, 羰基氧负离子形成的“氧负离子洞”通常被两个邻近氨基酸残基主链NH基团而稳定。这些NH基团通常来自活性中心Ser的下一位残基和位于 β 3折叠转角位置的残基。第二步亲核进攻反应与第一步相似, 常发生分子内基团成环释放、水分子进攻水解或更罕见的分子间基团进攻生成二元复合物。即外源亲核试剂(通常是水分子)或内源的亲核基团(通常为中间体结构中的羟基或氨基基团)被His活化, 随后亲核进攻(peptidyl)acyl-O-Ser复合物的C1位, 形成线性水解的产物、大环内酯化或大环内酰胺化的产物。TEII催化过程的化学反应过程与TEI类似, 但TEII的第二步亲核进攻反

应通常由水分子发起，主要以水解反应形式去除非反应性的（氨基）酰基残基或异常中间体来恢复

NRPSs 和 PKSs 的性能，以保障 PSK/NRPS 装配线的合成效率和提高终产物产量^[12, 43-44]。

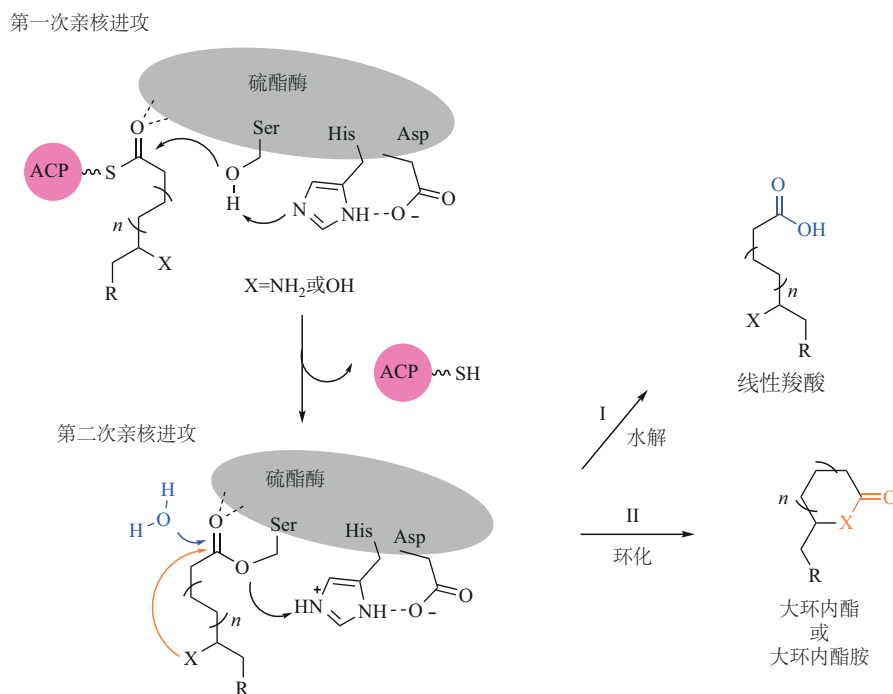


Fig. 5 Mechanism on NRPS/PKS chain release catalyzed by TEI^[42]

图5 TEI催化的NRPS和PKS链释放机制^[42]

以匹马霉素硫酯酶Pim TE 催化聚酮链释放为例^[24]，TE 催化循环通常历经以下几个过程：

2.1.1 底物上载（第一次亲核进攻）

经匹马霉素聚酮合酶完整延伸后的全长聚酮中间体加载在 ACP12 结构域的 Ppant 上。随着 ACP12 结构域锚定于 Pim TE 蛋白 N 端底物通道入口右侧，聚酮中间体在 Ppant 的牵引下进入 Pim TE 底物通道。Pim TE 底物通道入口含有大量极性氨基酸，通过与 Ppant 之间的相互作用，辅助底物深入底物通道，进行底物识别和环化构象预组织。

第一个碳四面体中间体形成。Pim TE 活性中催化三联体中的 Ser138 在底物上载时首先被 His261 去质子活化。Asp166 极性侧链可增强 His261 的拔质子能力。当底物和 Ppant 之间的硫酯键靠近 Ser138 时，去质子化的 Ser138 侧链羟基发起亲核进攻，与匹马霉素聚酮链 C1 形成碳氧键，形成第一个碳四面体中间体（图 6，iii）。在此过程中 C1 位羰基氧通常与活性中心 Ser138 下一位残基 Ser139 的主链 NH 基团形成氢键相互作用，稳定氧

负离子洞，促进亲核反应的发生。

酶-底物复合物形成和 ACP-Ppant 离去（图 6，iv）。碳四面体中间体形成后，His261 充当质子供体，将质子传递给 ACP-Ppant 末端的硫原子，完成硫酯键的断裂，形成酶-底物酯酰复合物，完成底物的上载过程。此外，ACP-Ppant 的离去发生在 TE 催化进程中的具体环节目前尚无文献报道，不排除 ACP-Ppant 在第二次亲核进攻结束后离去的可能性。

2.1.2 底物环化（第二次亲核进攻）

一般认为，在底物上载时 TE 对底物结构的容忍性较高，但在底物环化释放时存在一定的严谨性和底物选择性。底物在 TE 底物结合口袋中正确的环化构象预组织是 TE 催化底物环化释放的关键。一旦底物环化构象预组织就绪，后续的第二次亲核进攻随之发生，完成底物的环化。Pim TE 催化 26 元匹马霉素内酯环的形成是通过末端羟基 C25-OH 亲核进攻 C1 形成 C—O 键来实现的。

第二个碳四面体中间体形成。His261 在

Asp166的活化下, 辅助末端羟基C25-OH去质子化。活化后的末端羟基C25-O⁻亲核进攻C1原子, 形成C—O共价键, 形成第二个碳四面体中间体结构(图6, vii)。同样地, Ser139的主链NH与C1羰基的“氧负离子”形成氢键相互作用稳定该四面体构型。

2.1.3 环化产物的释放

环化产物被活性中心释放。第二个碳四面体中间体形成后, 之前质子化的H261-H⁺进将质子归还给Ser138, 使内酯环产物与活性中心(Ser138)脱离, 完成环化产物的释放, Ser138和His261恢复初始构象(图6, viii)。最后环化产物从C端退出底物通道, Pim TE完成一轮催化反应后, 空置的底物通道再次接收下一轮催化的底物。

理论上, Pim TE 活性中心完成环化后即代表一轮催化反应的结束, 但环化产物退出Pim TE 底物口袋才是真正意义上的产物释放。Pim TE 和环状产物共晶结构(PDB: 7VO5)正好捕捉了活性中心完成环化产物释放的瞬间, 说明这个瞬间结构是整个Pim TE 催化进程中最稳定的构象。共晶结构中, 环化产物的C1-羰基仍与Ser139侧链羟基形成氢键相互作用, 底物通道内Gln29通过一分子水与产物的C3-OH和C7-OH形成“水桥”结构。只有当这些环化产物与底物口袋之间的“羁绊”完全解除, 产物才能完整退出底物通道^[24]。这进一步暗示产物和底物口袋之间的“羁绊”可能是TE催化过程中的瓶颈之一^[24]。

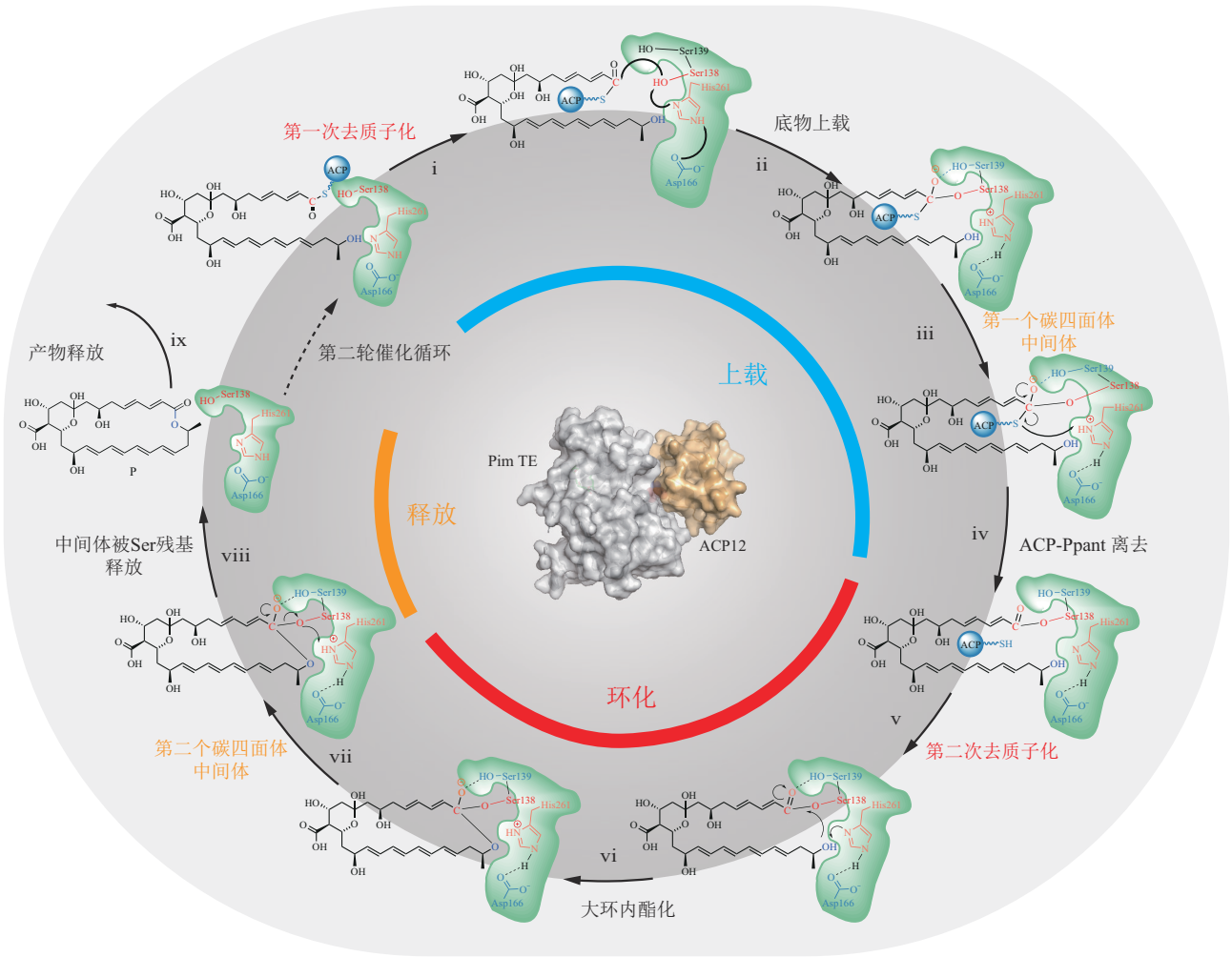


Fig. 6 Schematic diagram of Pim TE catalyzed macrolactonization and chain release of a polyketide intermediate^[24]

图6 Pim TE催化聚酮中间体环化释放反应示意图^[24]

2.2 TE催化机制的计算模拟研究

TEI常作为末端结构域出现在NRPS和PKS复合酶系中,其催化的底物通常为经完全延伸和修饰的成熟长链中间体,结构复杂且难以获得。而且,TE催化的链释放反应是一个典型的Bi-Bi反应模型,反应过程中存在酶-底物共价中间体结构,涉及两次亲核进攻和两个碳四面体结构。常规体外生化手段和蛋白质结构解析都存在较大的局限性,无法直接分析TE催化过程中质子转移和反应能量变化,无法识别反应过程中关键氨基酸的重要作用,更无法逐步定位每个反应中间体与酶的相互作用。因此在蛋白质结构解析的基础上,结合分子模拟和量化计算手段可以有效克服湿实验对TE催化机制研究的不足,被广泛应用于对TE环化和水解机制的深入研究。

2.2.1 “预反应态”模型解释DEBS TE和Pik TE对底物环化机制

Chen等^[45]对DEBS TE的计算模拟研究表明,PKS产物在TE大环化和释放过程中存在预反应态(pre-reaction state)-作用机制。采用分子动力学(molecular dynamics, MD)模拟结合量子力学(quantum mechanics, QM)以及量子力学与分子力学(molecular mechanics, MM)联用的QM/MM计算方法对DEBS TE催化4种不同结构的底物分子(图7a中1, 2, 3和天然底物)环化或水解释放反应进行研究。Chen等^[45]通过能量计算优化到一个具有11.6 kcal/mol能垒的关键预反应态。在MD模拟中,该预反应态结构能稳定存在一定时间,给上载底物的末端羟基氧攻击C1-羰基碳创造机会。然而,对于底物2/3(图7a)(C7-羰基被羟基取代),在类似的MD模拟中几乎没有观察到这种环化预反应态的形成,且底物2/3的尾部倾向朝酶活性口袋出口位置运动,导致水分子进入活性中心,促使水解反应发生。对于DEBS TE的天然底物,MD模拟中发现其底物可以与TE形成稳定的预反应态结构,其占比明显优于底物1和2/3(图7a),说明预反应态模型可以用来评估TE催化环化反应的可能性(图7b)。

苦霉素PKS复合酶系在延伸过程中,存在非常独特的2种链释放模式^[46],即苦霉素PKS复合酶系可以同时催化12元大环内酯10-deoxymethynolide(10-DML)和14元大环内酯narbonolide的环化释放。研究发现,10-DML的环化释放是因为PikAIII的ACP结构域中加载的中间

体hexaketide(图7a中1)由于三维空间结构靠近PikAV末端的TEI,而被TEI直接催化释放。Shi等^[47]同样通过MD结合QM/MM计算模拟分析揭示了Pik TE预反应态的结构特征和关键亲疏水相互作用。相较于narbonolide的前体heptaketide(图7a中4),Pik TE与10-DML前体hexaketide(图7a中1)更倾向于形成稳定的预反应状态,这对环化释放至关重要。此外,Shi等^[47]还计算了反应势能面来研究Pik TE催化10-DML环化的分子机制,揭示Pik TE-底物1反应体系的 re -face亲核进攻的能垒比Pik TE-4体系低近4.1 kcal/mol,与实验结果一致(10-DML的产率优于narbonolide)。

2.2.2 分子动力学模拟结合量化计算揭示Pik TE和TE_{S148C}对底物的立体选择性催化机制

Hansen等^[9]的研究指出,Pik TE对非天然底物的催化存在瓶颈效应,是限制大环内酯化产物产率的关键因素。Koch等^[48]通过MD模拟揭示了Pik TE对非天然异构化底物的催化障碍的分子基础,即非天然底物5(图7a)的固有构象导致其不利于被Pik TE环化释放。量化计算结果表明,TE_{WT}和TE_{S148C}催化天然底物1(图7a)环化(18.2 kcal/mol和14.8 kcal/mol)比非天然异构化的底物5(图7a)(22.4 kcal/mol和16.0 kcal/mol)更具能量优势,且TE_{S148C}突变有效降低了天然底物和非天然异构底物的反应能垒^[48]。Pik TE活性中心Ser突变为Cys,显著提高其催化大环化的能力,为高效合成具有生物活性的大环内酯类化合物提供了新的TE理性设计策略^[49]。

2.2.3 多烯硫酯酶理论计算揭示两性大环分子环化机制并指导硫酯酶定向改造

匹马霉素、两性霉素、制霉菌素等多烯抗生素是典型的两性聚酮类活性天然产物,因结构中存在特殊的共轭多烯结构和良好的抗真菌活性而被关注。多烯类抗生素内酯环尺寸偏大,一般为20~44元大环内酯,其内酯环一侧是由3~8个共轭双键组成的疏水共轭平面,另一侧为多个羟基结构形成的多元醇亲水区。Zhou等^[24]以Pim TE和环化产物的复合物结构(PDB: 7VO5)为基础上搭建了Pim TE催化天然羧基底物和非天然脱羧-甲基底物的MD模拟模型,轨迹聚类显示天然底物的环化优势构象占比显著高于非天然脱羧-甲基底物。天然底物和Pim TE底物通道可以保持平衡且稳定持续的氢键互作网络,促进预反应态的生成和成环释放。而非天然脱羧-甲基底物因缺乏肘部的氢键平

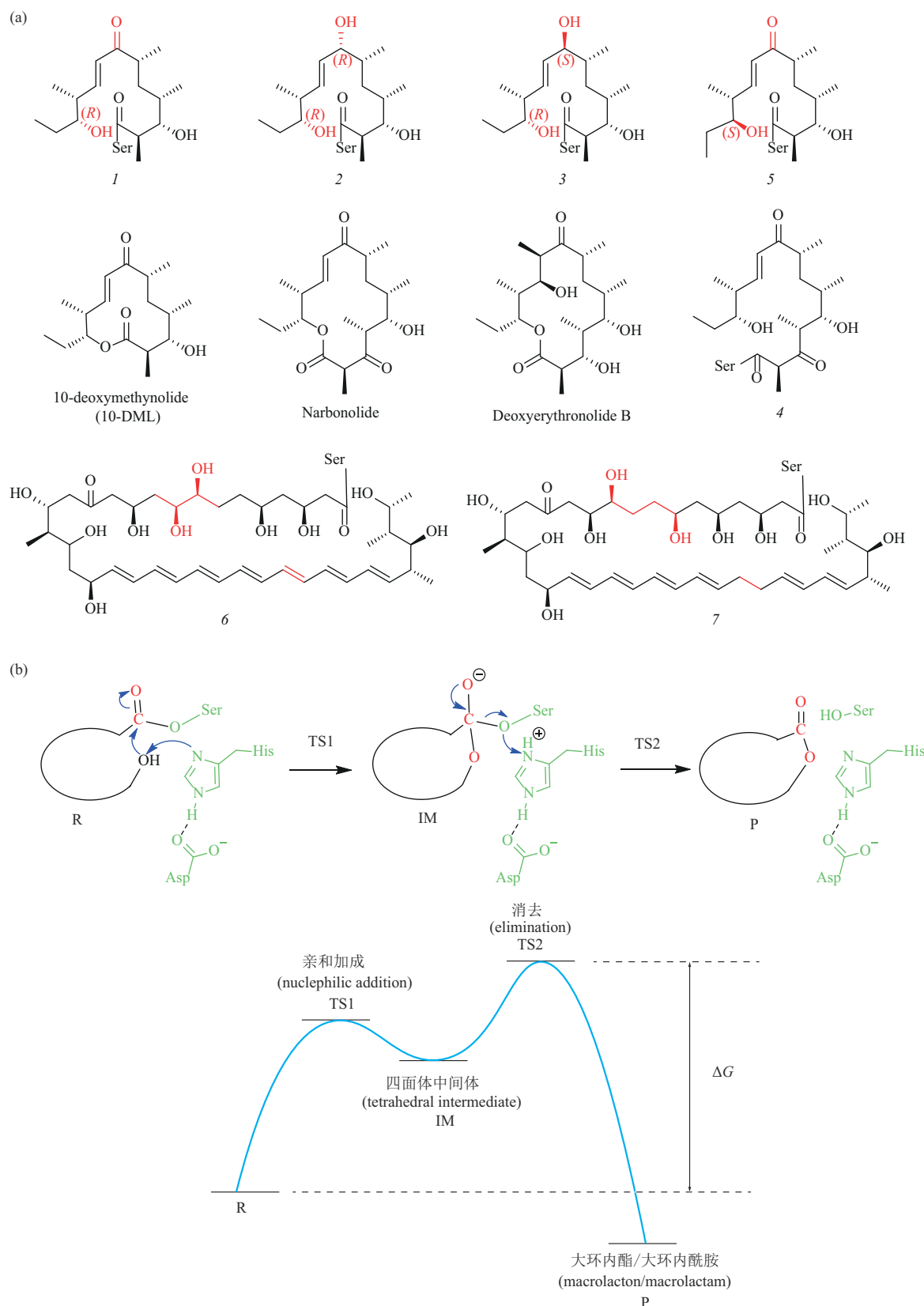


Fig. 7 Chemical structures of the substrates or products related in TE catalysis and the schematic diagram of TE reaction states calculated by QM/MM analysis [45, 47–48]

图7 硫酯酶催化涉及的部分化合物分子结构和量化计算分析的硫酯酶反应示意图 [45, 47–48]

(a) 本文提及的硫酯酶催化相关化合物分子结构。10-deoxymethynolide (10-DML): 脱氧酒毒素内酯前体; narbonolide: 苦霉素内酯前体; deoxyerythronolide B: 脱氧红霉素内酯B。(b) 硫酯酶反应及其能量变化示意图。R: 预反应态 (prereaction state); TS1: 过渡态1 (transition state 1); IM: 中间态 (intermediate); TS2: 过渡态2 (transition state 2); P: 终产物 (product); ΔG : 反应活化能。

衡，末端羟基无法稳定在亲核进攻的有效距离内，导致环化优势构象（预反应态）占比极低。Zhou等^[24]进一步通过构建Pim TE蛋白突变体文库，结合体外生化活性分析，筛选了对非天然脱羧-甲基中间体催化效率显著提升的突变靶点，并应用于体内遗传改造，成功提升了低毒脱羧-甲基匹马霉素衍生物的产量。此外，部分Pim TE突变位点MD模拟轨迹也证实了其对底物识别、环化释放等过程

的有利影响^[50]。Jiang等^[51]从计算生物学角度分析了匹马霉素等多烯类抗生素聚酮骨架结构中的“半缩酮”环的形成机制，并推测Pim TE可能是双功能酶，协同6元半缩酮环和26元大环内酯的串联形成。Wang等^[52]还搭建了AMB TE和NYS TE分别催化44元两性霉素大环内酯和制霉菌素大环内酯环化释放的4个MD模拟体系，发现NYS TE对其天然底物Nys（图7a中7）的识别特异性高于Amb，而AMB TE对两种底物（图7a中6和7）均有良好的耐受性。虽然两性霉素和制霉菌素聚酮内酯环仅存在一个羟基位置和一个双键的结构差异，但QM/MM计算显示两种TE对天然底物的环化都比交换底物更有利，AMB-TE-Amb和AMB-TE-Nys的环化能垒分别为14.0 kcal/mol和22.7 kcal/mol，而NYS-TE-Nys和NYS-TE-Amb的能垒分别为17.5和25.7 kcal/mol（表2）。

Table 2 Relative energetics calculated in the TE-catalyzed macrolactonization by QM/MM method
表2 QM/MM方法计算的硫酯酶催化大环内酯化反应的能量变化

硫酯酶	化合物编号 (图7a)	原参考文献中 的化合物编号	反应自由能/(kcal·mol ⁻¹)					反应活化能 ΔG/(kcal·mol ⁻¹)	参考 文献
			R	TS1	IM	TS2	P		
DEBS TE	1	30	0.0	9.9	9.9	10.2	-17.6	9.9	[45]
DEBS TE	2	31	0.0	10.9	10.2	11.0	-17.8	10.9	[45]
DEBS TE	3	32	0.0	9.1	10.0	10.1	-19.7	9.1	[45]
Pik TE	1/re-face	1	0.0	10.2	9.4	16.3	-1.4	16.3	[47]
Pik TE	1/si-face	1	0.0	12.3	11.3~13.7	19.9	0.0	19.9	[47]
Pik TE	4	2	0.0	15.4	15.3	20.4	-7.7	20.4	[47]
Pik TE/WT	1	13	0.0	19.1	20.8	22.4	-3.3	22.4	[48]
Pik TE/WT	5	14	0.0	14.2	12.7	18.2	-9.8	18.2	[48]
Pik TE/S148C	1	13	0.0	酰基协同取代: 16.0			-9.2	16.0	[48]
Pik TE/S148C	5	14	0.0	酰基协同取代: 14.8			-12.9	14.8	[48]
AMB TE	6	Amb	0.0	10.0	9.1	14.0	2.3	14.0	[52]
AMB TE	7	Nys	0.0	14.6	10.3	22.7	12.5	22.7	[52]
NYS TE	6	Amb	0.0	18.9	17.3	25.7	10.1	25.7	[52]
NYS TE	7	Nys	0.0	17.5	4.1	14.1	-1.7	17.5	[52]

此外，MD模拟赋能QM/MM能垒分析的计算生物手段，还被应用于一些具有特殊催化性能的NRPS TEI的研究中。双功能硫酯酶NocTE是NRPS NocA-NocB组装线上D-构型产物的“守门人”，Yu等^[53]通过经典的MD模拟、MM-PBSA（molecular mechanics poisson-Boltzmann surface area）计算和限制性MD和拉伸MD模拟阐明了三肽/五肽长度和β内酰胺环在Noc TE底物选择性中的重要作用，并通过定位关键残基His1808发现其与非对映体*epi*-nocardicin G易形成稳定氢键而阻碍环化产物释放，揭示Noc TE对产物立体构型选择性释放的分子基础。Kobayashi等^[54-55]搭建了NRPS青霉素结合蛋白（penicillin-binding protein, PBP）型硫酯酶SurE的计算模型研究SurE对底物

的识别和催化过程，发现在底物识别过程中，底物C端D-Leu可以通过氧负离子洞而稳定，N端L-Ile被由Tyr154-Lys66-Asn156等残基形成的氢键互作网络识别，结合Arg446对底物的锚定作用，共同促进了SurE预反应态结构的形成，并通过翻越约19.4 kcal/mol的环化能垒完成对大环内酰胺化产物分子的释放。上述对SurE的底物识别和催化特点的计算模拟还进一步指导了4种新型肽前体的设计，并通过SurE有效地环化成环肽^[55]。计算生物学研究方法和理论的发展^[56-57]，不仅有效辅助TE底物识别和催化机制的深入剖析，更为TE理性设计提供了设计思路和可行的突变策略，还将促进新型聚酮类和非核糖体肽类活性天然产物的结构改造和产量提升研究^[58]。

3 总结与展望

聚酮类和非核糖体肽类活性天然产物是宝贵的临床药物先导化合物资源库^[1, 3, 59-61], TE作为这些天然产物最重要的链释放途径之一, 其蛋白质结构、催化机制、底物选择性等一直备受关注^[11, 62]。但相关TE结构和催化机制的研究还存在以下局限性: 首先, 目前已解析报道的TE大多是*apo* TE结构, 或者少数TE和环化产物的复合物结构, 但与天然底物的*holo* TE构象, 尤其是ACP-底物-TE三元复合物结构研究报道非常有限, 这将直接阻碍破解TE识别底物的分子奥秘, 限制计算模拟的精准性; 其次, PKS和NRPS合酶作为多模块多结构域的复合酶系, 其结构域和模块之间的空间构象组织对催化活性也同样关键^[12, 63-64], 目前很少含TEI的PKS或NRPS复合酶结构被解析报道^[65], 揭秘TEI在复合酶系的构象组织有望从结构生物学的角度剖析Pik TE等催化多个底物环化释放的分子机制; 此外, 受限于计算方法理论和算力, 更大尺度的MD模拟, 比如ACP/PCP-底物-TE复合体系, 乃至整个PKS或NRPS复合酶系, 以及全蛋白质层面的QM计算都尚未应用于TE研究, 相关研究的开展有助于发现底物结合口袋和催化活性中心以外相对远端的残基对TE催化的间接影响, 更好地指导TE理性设计。在未来, 通过TE定向改造释放其天然底物^[24], 并进一步挂在到ACP/PCP结构域中, 与硫酯酶无活性或低活性突变体共孵育, 可能可以捕获ACP-底物-TE三元复合物结构。不仅如此, 通过PKS或NRPS末端模块的共表达和纯化, 并尝试挂载天然底物, 结合冷冻电镜技术, 有望揭开天然产物链释放过程中TE和ACP/PCP等关键结构域的构象组织的奥秘。

将来, 通过合成生物学工具的应用, 针对性地重编程模块化的PKS和NRPS来生产特定的、高价值的天然产物衍生物是重要发展方向^[4, 5, 66-67]。但是, 生物合成途径上游中间体的结构改变, 往往引起下游催化酶效率的下降, 致使新结构、新活性衍生物的产量很低, 难以满足后续的临床检定和大规模生产的需求。而TE在这些活性非天然产物生物合成的过程中发挥着重要的底物选择功能, 是天然产物及其衍生物合成的关键催化酶^[62]。阐明TE催化的结构生物学机制, 并在PKS/NRPS重编程的基础上偶联TE的定向设计与改造, 可改善TE对新结构底物的催化效率, 并大量积累目标产物。以TE

为切入点, 以蛋白质结构和催化机制为导向的TE理性设计和定向改造将为生产新的天然产物类似物提供更灵活的催化。

参 考 文 献

- [1] Sang M, Feng P, Chi L P, *et al.* The biosynthetic logic and enzymatic machinery of approved fungi-derived pharmaceuticals and agricultural biopesticides. *Nat Prod Rep*, 2024, **41**(4): 565-603
- [2] Wang M, Li H, Li J, *et al.* Streptomyces strains and their metabolites for biocontrol of phytopathogens in agriculture. *J Agric Food Chem*, 2024, **72**(4): 2077-2088
- [3] Wang L, Lu H, Jiang Y. Natural polyketides act as promising antifungal agents. *Biomolecules*, 2023, **13**(11): 1572
- [4] Bozhüyük K A J, Präve L, Kegler C, *et al.* Evolution-inspired engineering of nonribosomal peptide synthetases. *Science*, 2024, **383**(6689): eadg4320
- [5] Mabesoone M F J, Leopold-Messer S, Minas H A, *et al.* Evolution-guided engineering of trans-acyltransferase polyketide synthases. *Science*, 2024, **383**(6689): 1312-1317
- [6] Yang M, Hao Y, Liu G, *et al.* Enhancement of acyl-CoA precursor supply for increased avermectin B1a production by engineering meilingmycin polyketide synthase and key primary metabolic pathway genes. *Microb Biotechnol*, 2024, **17**(5): e14470
- [7] Zhai G, Zhu Y, Sun G, *et al.* Insights into azalomycin F assembly-line contribute to evolution-guided polyketide synthase engineering and identification of intermodular recognition. *Nat Commun*, 2023, **14**(1): 612
- [8] Gehret J J, Gu L, Gerwick W H, *et al.* Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J Biol Chem*, 2011, **286**(16): 14445-14454
- [9] Hansen D A, Koch A A, Sherman D H. Identification of a thioesterase bottleneck in the pikromycin pathway through full-module processing of unnatural pentaketides. *J Am Chem Soc*, 2017, **139**(38): 13450-13455
- [10] Yu Y, van der Donk W A. PEARL-catalyzed peptide bond formation after chain reversal by ureido-forming condensation domains. *ACS Cent Sci*, 2024, **10**(6): 1242-1250
- [11] Fraley A E, Dieterich C L, Mabesoone M F J, *et al.* Structure of a promiscuous thioesterase domain responsible for branching acylation in polyketide biosynthesis. *Angew Chem Int Ed*, 2022, **61**(39): e202206385
- [12] Wang S, Brittain W D G, Zhang Q, *et al.* Aminoacyl chain translocation catalysed by a type II thioesterase domain in an unusual non-ribosomal peptide synthetase. *Nat Commun*, 2022, **13**(1): 62
- [13] Pourmasoumi F, De S, Peng H, *et al.* Proof-reading thioesterase boosts activity of engineered nonribosomal peptide synthetase. *ACS Chem Biol*, 2022, **17**(9): 2382-2388
- [14] Geyer K, Hartmann S, Singh R R, *et al.* Multiple functions of the type II thioesterase associated with the phoslactomycin polyketide

- synthase. *Biochemistry*, 2022, **61**(23): 2662-2671
- [15] Little R, Trottmann F, Preissler M, *et al.* An intramolecular thioesterase domain catalyses chain release in the biosynthesis of a cytotoxic virulence factor. *RSC Chem Biol*, 2022, **3**(9): 1121-1128
- [16] Liu T, You D, Valenzano C, *et al.* Identification of NanE as the thioesterase for polyether chain release in nanchangmycin biosynthesis. *Chem Biol*, 2006, **13**(9): 945-955
- [17] Liu T, Lin X, Zhou X, *et al.* Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin. *Chem Biol*, 2008, **15**(5): 449-458
- [18] Swarbrick C M D, Nanson J D, Patterson E I, *et al.* Structure, function, and regulation of thioesterases. *Prog Lipid Res*, 2020, **79**: 101036
- [19] Horsman M E, Hari T P A, Boddy C N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate?. *Nat Prod Rep*, 2016, **33**(2): 183-202
- [20] Kim M S, Bae M, Jung Y E, *et al.* Unprecedented noncanonical features of the nonlinear nonribosomal peptide synthetase assembly line for WS9326A biosynthesis. *Angew Chem Int Ed*, 2021, **60**(36): 19766-19773
- [21] Tsai S C, Miercke L J, Krucinski J, *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci USA*, 2001, **98**(26): 14808-14813
- [22] Tsai S C, Lu H, Cane D E, *et al.* Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry*, 2002, **41**(42): 12598-12606
- [23] Bruner S D, Weber T, Kohli R M, *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure*, 2002, **10**(3): 301-310
- [24] Zhou Y, Tao W, Qi Z, *et al.* Structural and mechanistic insights into chain release of the polyene PKS thioesterase domain. *ACS Catal*, 2022, **12**(1): 762-776
- [25] Scaglione J B, Akey D L, Sullivan R, *et al.* Biochemical and structural characterization of the tautomycin thioesterase: analysis of a stereoselective polyketide hydrolase. *Angew Chem Int Ed*, 2010, **49**(33): 5726-5730
- [26] Akey D L, Kittendorf J D, Giraldez J W, *et al.* Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol*, 2006, **2**(10): 537-542
- [27] Samel S A, Wagner B, Marahiel M A, *et al.* The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol*, 2006, **359**(4): 876-889
- [28] Patel K D, D'Andrea F B, Gaudelli N M, *et al.* Structure of a bound peptide phosphonate reveals the mechanism of nocardicin bifunctional thioesterase epimerase-hydrolase half-reactions. *Nat Commun*, 2019, **10**(1): 3868
- [29] Huguenin-Dezot N, Alonzo D A, Heberlig G W, *et al.* Trapping biosynthetic acyl-enzyme intermediates with encoded 2, 3-diaminopropionic acid. *Nature*, 2019, **565**(7737): 112-117
- [30] Liu Y, Zheng T, Bruner S D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol*, 2011, **18**(11): 1482-1488
- [31] Frueh D P, Arthanari H, Koglin A, *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature*, 2008, **454**(7206): 903-906
- [32] Korman T P, Crawford J M, Labonte J W, *et al.* Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci USA*, 2010, **107**(14): 6246-6251
- [33] Guntaka N S, Healy A R, Crawford J M, *et al.* Structure and functional analysis of ClbQ, an unusual intermediate-releasing thioesterase from the colibactin biosynthetic pathway. *ACS Chem Biol*, 2017, **12**(10): 2598-2608
- [34] Koglin A, Löhr F, Bernhard F, *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature*, 2008, **454**(7206): 907-911
- [35] Whicher J R, Florova G, Sydor P K, *et al.* Structure and function of the RedJ protein, a thioesterase from the prodigine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem*, 2011, **286**(25): 22558-22569
- [36] Curran S C, Pereira J H, Baluyot M J, *et al.* Structure and function of BorB, the type II thioesterase from the borrelidin biosynthetic gene cluster. *Biochemistry*, 2020, **59**(16): 1630-1639
- [37] Badding E D, Grove T L, Gadsby L K, *et al.* Rerouting the pathway for the biosynthesis of the side ring system of nosiheptide: the roles of NosI, NosJ, and NosK. *J Am Chem Soc*, 2017, **139**(16): 5896-5905
- [38] Claxton H B, Akey D L, Silver M K, *et al.* Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *J Biol Chem*, 2009, **284**(8): 5021-5029
- [39] Ohlemacher S I, Xu Y, Kober D L, *et al.* YbtT is a low-specificity type II thioesterase that maintains production of the metallophore yersiniabactin in pathogenic enterobacteria. *J Biol Chem*, 2018, **293**(51): 19572-19585
- [40] Galea C A, Roberts K D, Zhu Y, *et al.* Functional characterization of the unique terminal thioesterase domain from polymyxin synthetase. *Biochemistry*, 2017, **56**(4): 657-668
- [41] Kreitler D F, Gemmell E M, Schaffer J E, *et al.* The structural basis of N-acyl- α -amino- β -lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat Commun*, 2019, **10**(1): 3432
- [42] Little R F, Hertweck C. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. *Nat Prod Rep*, 2022, **39**(1): 163-205
- [43] Kotowska M, Pawlik K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl Microbiol Biotechnol*, 2014, **98**(18): 7735-7746
- [44] Flatt P M, O'Connell S J, McPhail K L, *et al.* Characterization of the initial enzymatic steps of barbamide biosynthesis. *J Nat Prod*, 2006, **69**(6): 938-944
- [45] Chen X P, Shi T, Wang X L, *et al.* Theoretical studies on the

- mechanism of thioesterase-catalyzed macrocyclization in erythromycin biosynthesis. *ACS Catal*, 2016, **6**(7): 4369-4378
- [46] Kittendorf J D, Beck B J, Buchholz T J, *et al.* Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chem Biol*, 2007, **14**(8): 944-954
- [47] Shi T, Liu L, Tao W, *et al.* Theoretical studies on the catalytic mechanism and substrate diversity for macrocyclization of pikromycin thioesterase. *ACS Catal*, 2018, **8**(5): 4323-4332
- [48] Koch A A, Hansen D A, Shende V V, *et al.* A single active site mutation in the pikromycin thioesterase generates a more effective macrocyclization catalyst. *J Am Chem Soc*, 2017, **139**(38): 13456-13465
- [49] Qiao L, Fang J, Zhu P, *et al.* A novel chemoenzymatic approach to produce cilengitide using the thioesterase domain from *Microcystis aeruginosa* microcystin synthetase C. *Protein J*, 2019, **38**(6): 658-666
- [50] Fan S, Wang R, Li C, *et al.* Insight into structural characteristics of protein-substrate interaction in pimaricin thioesterase. *Int J Mol Sci*, 2019, **20**(4): 877
- [51] Jiang C, Zhou Y, Tao W, *et al.* Theoretical studies of mutual effects between 6-m-r hemiketalization and 26-m-r lactonization in pimaricin thioesterase. *Chem Asian J*, 2023, **18**(7): e202201229
- [52] Wang R, Tao W, Liu L, *et al.* Insights into specificity and catalytic mechanism of amphotericin B/nystatin thioesterase. *Proteins*, 2021, **89**(5): 558-568
- [53] Yu Q, Xie L, Li Y, *et al.* Exploring the molecular basis of substrate and product selectivities of nocardicin bifunctional thioesterase. *Interdiscip Sci*, 2022, **14**(1): 233-244
- [54] Kobayashi M, Fujita K, Matsuda K, *et al.* Streamlined chemoenzymatic synthesis of cyclic peptides by non-ribosomal peptide cyclases. *J Am Chem Soc*, 2023, **145**(6): 3270-3275
- [55] Du Z, Ma Y, Shen Y, *et al.* Exploring the substrate stereoselectivity and catalytic mechanism of nonribosomal peptide macrocyclization in surugamides biosynthesis. *iScience*, 2024, **27**(2): 108876
- [56] Tzeliou C E, Mermigki M A, Tzeli D. Review on the QM/MM methodologies and their application to metalloproteins. *Molecules*, 2022, **27**(9): 2660
- [57] Hollingsworth S A, Dror R O. Molecular dynamics simulation for all. *Neuron*, 2018, **99**(6): 1129-1143
- [58] Nava A A, Roberts J, Haushalter R W, *et al.* Module-based polyketide synthase engineering for *de novo* polyketide biosynthesis. *ACS Synth Biol*, 2023, **12**(11): 3148-3155
- [59] Avalon N E, Reis M A, Thornburg C C, *et al.* Leptochelins A-C, cytotoxic metallophores produced by geographically dispersed leptothoe strains of marine cyanobacteria. *J Am Chem Soc*, 2024, **146**(27): 18626-18638
- [60] Zang H, Cheng Y, Li M, *et al.* Mutagenetic analysis of the biosynthetic pathway of tetramate bripiodionen bearing 3-(2H-pyran-2-ylidene)pyrrolidine-2, 4-dione skeleton. *Microb Cell Fact*, 2024, **23**(1): 87
- [61] Yi J S, Kim J M, Ban Y H, *et al.* Modular polyketide synthase-derived insecticidal agents: from biosynthesis and metabolic engineering to combinatorial biosynthesis for their production. *Nat Prod Rep*, 2023, **40**(5): 972-987
- [62] Su L, Hôtel L, Paris C, *et al.* Engineering the stambomycin modular polyketide synthase yields 37-membered mini-stambomycins. *Nat Commun*, 2022, **13**(1): 515
- [63] Koglin A, Mofid M R, Löhr F, *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science*, 2006, **312**(5771): 273-276
- [64] Bagde S R, Mathews I I, Fromme J C, *et al.* Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science*, 2021, **374**(6568): 723-729
- [65] Tanovic A, Samel S A, Essen L O, *et al.* Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science*, 2008, **321**(5889): 659-663
- [66] Soohoo A M, Cogan D P, Brodsky K L, *et al.* Structure and mechanisms of assembly-line polyketide synthases. *Annu Rev Biochem*, 2024, **93**(1): 471-498
- [67] Whicher J R, Dutta S, Hansen D A, *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*, 2014, **510**(7506): 560-564

Structural Elucidation and Catalytic Mechanisms of PKS and NRPS Thioesterase Domains*

ZHOU Yu-Cong^{1,2)**}, SHI Ting²⁾, LU Chen-Yang³⁾, LIU Hao⁴⁾, BAI Lin-Quan^{2)**}

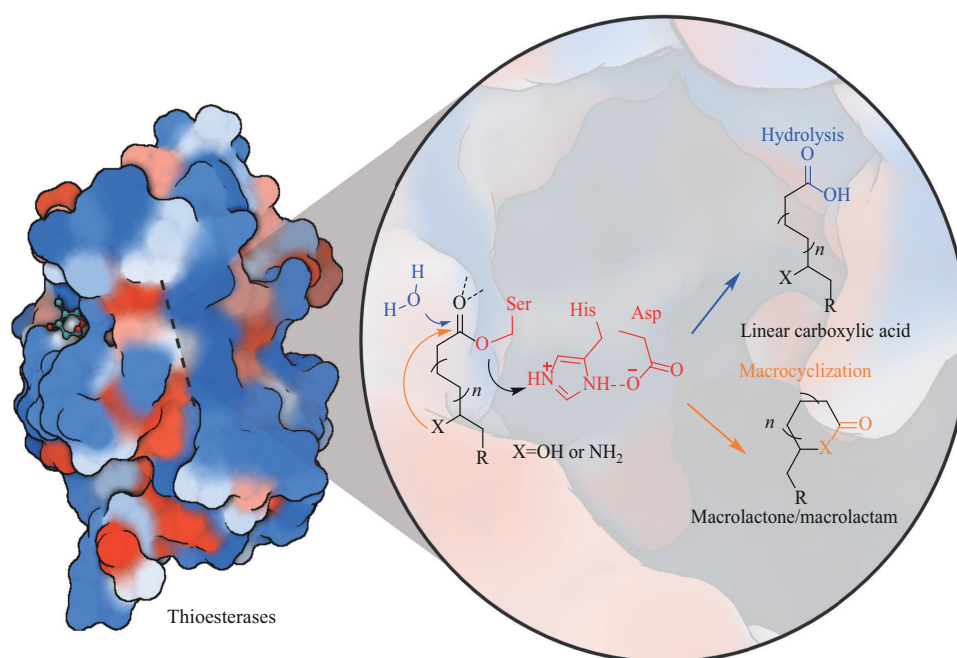
⁽¹⁾College of Biological and Environmental Science, Zhejiang Wanli University, Ningbo 315100, China;

⁽²⁾School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China;

⁽³⁾School of Marine Science, Ningbo University, Ningbo 315832, China;

⁽⁴⁾School of Basic Medical Science, School of Medicine, Ningbo University, Ningbo 315211, China)

Graphical abstract



Abstract Polyketides (PKs) and non-ribosomal peptides are the most important drug-leads for human, animal, and plant diseases. The conserved modular architectures and biosynthetic assembly line of polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS) endow PKs and NRPs with extremely diverse structures and activities and bring infinite possibilities to edit and modify the backbone structure of PKs and NRPs by adding, removing, inactivating and replacing PKS/NRPS modules or domains. The biosynthetic machinery of microbial

* This work was supported by grants from China Postdoctoral Science Foundation (2023M743263) and Zhejiang Provincial Postdoctoral Scholarship (ZJ2023065).

** Corresponding author.

ZHOU Yu-Cong. Tel: 86-18817555727, E-mail: zhouyucong@alumni.sjtu.edu.cn

BAI Lin-Quan. Tel: 86-21-62932418, E-mail: bailq@sjtu.edu.cn

Received: July 1, 2024 Accepted: October 14, 2024

polyketide natural products has evolved delicately with specific recognition and efficient catalysis of upstream intermediates by downstream enzymes/domains. Therefore, manipulations of PKS/NRPS and their related tailoring enzymes usually lead to attenuated production or abolished accumulation of intermediates with modified structures. As the terminal domain of most PKS and NRPS, thioesterases (TEs) play crucial roles in substrate selection during the chain release of these bioactive natural products, serving as pivotal bottleneck steps in their late-stage biosynthesis. TEs mainly perform chain hydrolysis or ester transfer reactions by nucleophilic attack of foreign nucleophiles such as H₂O. Meanwhile, TEs also undergo nucleophilic attack by intramolecular oxygen atom, nitrogen atom, or carbon atom to achieve macrolactonization, macrolactamization, or Claisen condensation, respectively. There are two main classes of TEs involved in natural product biosynthesis. Type I TEs (TEIs) are commonly found in type I *cis*-AT PKS, *trans*-AT PKS, NRPS, and fungal PKS/NRPS, which are mainly located at the end module of synthase. In addition to TEIs, there is also a class of free type II TE (TEIIs), which catalyzes the release of incomplete or incorrectly extended intermediates during PKs and NRPs biosynthesis. Besides, a distinct class of free TE was identified in the chain release of polyether backbones, such as monensin and nanchangmycin. Since 2001, more than 20 crystal structures of TEs from diverse PKSs and NRPSs have been solved. The structural elucidation of TEs has unlocked the mystery of their structural and functional interaction, laid the foundation for the TE classification and mechanistic insight into the substrate selectivity and catalytic efficiency of TE, which further promotes the understanding of the chain release mechanism of natural products and better served the rational design of TE. Previous articles have systematically reviewed the structure, function, and regulatory mechanism of different TE families. Horsman *et al.* also reviewed the diversity, structure, and mechanism of TEs in PKSs and NRPSs. They put forward an insightful view that TEs might act as logic gates for substrate loading and chain releasing during the biosynthesis of natural products. It provides an important perspective for studying the evolution and functional prediction of TEs. This review summarizes the structural characteristics of various TE, focusing on the structural consistency of thioesterase to the catalytic mechanism. Additionally, this review follows the progress and limitations on the catalytic mechanism and computational simulation of type I TE, providing a detailed analysis of the chemical essence of thioesterase-catalyzed chain release reactions. This review aims to deliver revealing suggestions for the structural elucidation and mechanistic insights of TE, as well as its rational design for improved chain release of unnatural products.

Key words polyketide synthase, non-ribosomal peptide synthetase, thioesterase, protein structure, catalytic mechanism

DOI: 10.16476/j.pibb.2024.0288

CSTR: 32369.14.pibb.20240288